AWARD NUMBER:
W81XWH-12-1-0107

TITLE:
Use of a Novel Embryonic Mammary Stem Cell Gene Signature to Improve Human Breast Cancer Diagnostics and Therapeutic Decision Making

PRINCIPAL INVESTIGATOR:
Charles M. Perou

CONTRACTING ORGANIZATION:
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599

REPORT DATE:

December 2015

TYPE OF REPORT:

Final Report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| December 2015 | Final | 30 Sep 2012 - 29 Sep 2015 |

**4. TITLE AND SUBTITLE**

Use of a Novel Embryonic Mammary Stem Cell Gene Signature to Improve Human Breast Cancer Diagnostics and Therapeutic Decision Making

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
W81XWH-12-1-0107

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Wahl, Geoffrey M.; Perou, Charles; Spike, Benjamin; Lasken, Roger

E-Mail: cperou@med.unc.edu

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

The University of North Carolina at Chapel Hill
Lineberger Comprehensive Cancer Center
450 West Drive, CB# 7295
Chapel Hill NC 27599-7295

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

Approved for Public Release; Distribution Unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

Our major goals are to determine whether Fetal Mammary Stem Cell (fMaSC) signatures correlate with response to chemotherapy and metastasis in different breast cancer intrinsic subtypes (AIM1), and to develop single cell sequencing to produce highly refined fMaSC signatures (AIM2). Accomplishing these aims will enable us to: 1) better categorize distinct cell types within the fMaSC population, 2) identify biomarkers for prospective stem cell purification and in situ localization, and 3) identify candidate stem cell regulatory pathways that should reveal therapeutic targets and improved prognosticators and response biomarkers. In the most recent funding period, our bioinformatic analysis identified subsets of fMaSC signature genes that are coordinately expressed in archived human breast cancer gene expression data sets and assessed their prognostic and/or predictive power. We have thus far identified one subset exhibiting significant prognostic value distinct from existing and commonly used clinical variables in the preliminary data sets we have analyzed. We have also adapted a new microfluidics-based, single-cell capture and library preparation system to improve reproducibility in the generation of gene expression profiles from individual fMaSC. These advances provide proof of the principles underlying this grant and leave us well positioned to achieve its aims.

**15. SUBJECT TERMS**

Breast Cancer Prognosis, Mammary Stem Cells, Embryonic Development, Single Cell Transcriptomics

| 16. SECURITY CLASSIFICATION OF: Unclassified | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | 19b. TELEPHONE NUMBER *(include area code)* |
| U | U | U | UU | 10 | USAMRMC |

**Attachment 5. Statement of Work:** (W=Wahl Lab; P=Perou Lab; L=Lasken Lab)

**Task 1. Embryonic mammary stem cell signature refinement to delineate fMaSC traits in human cancers and to identify new targets for cancer stem cell directed therapies**

1a. The Perou lab will obtain the gene expression raw data of the fMaSC and fStroma samples previously characterized by the Wahl lab. First, using the current fMaSC signature of 600 genes, a score for each gene will be assigned based on its differential expression across the two groups. Second, using a cross validation approach, a genomic predictor will be created using the smallest gene list possible that can correctly discriminate the fMaSC vs. fStroma samples. Finally, samples used for the identification of the minimum gene list will be re-run onto the Fluidigm BioMark platform, and the optimal classification ability of the new genomic predictor will be re-tested. **P,W** (months l-2).

>>>This has been accomplished (see data below under section 1d).

1b. The latest and most extensive cell line microarray database of the Perou Lab will be used for this analysis. This data set includes 40 breast cancer cell lines, 12 human mammary epithelial and fibroblast cell lines (primary and immortalized), 3 human embryonic stem cell lines and 3 mesenchymal stem cell lines. For each cell line, the genomic Euclidian distance to the fMaSC and fStromal centroids will be calculated, and the ratio of both distances will be the final "enrichment score". **P** (months 2-4)

>>>This has been accomplished and gave an unexpected result. Namely we used the set of ~100 cell line genomic profiles coming from Prat et al., BCRT 2013 (PMID:24162158) and applied the fMaSC signature as implemented in Spike et al., and as implemented in Pfefferle et al. 2015; the analysis identified two important findings. First these two implementation of the same signature were highly concordant with each other (0.94 correlation), therefore this is a robust and reproducible signature. Second, the fMaSC signature was the most highly expressed in luminal breast cell lines, especially those that were HER2+ (i.e. BT474 and SKBR3). These findings are surprising because in vivo, the fMaSC signature is the most highly expressed in Basal-like breast cancer, while in vitro, it appears to be the most highly expressed in two Luminal & HER2+ cell lines, and not highly expressed in Basal-like cell lines (although we do note these genes are expressed in these Basal-like samples, just not as highly). With this result, this Aim has been accomplished and we now have at least 2 cell line models identified for the investigation of the fMaSC gene set *in vitro*.

1c. In addition, to further compare the levels of gene expression of the fMaSC- and fStromal-enriched populations with the Perou lab's cell lines, -12 fMaSC/fStromal samples will be collected, RNA isolated, amplified and hybridized onto the Perou Lab Whole Genome Custom Array Platform, and their gene expression profiles compared to the rest of cell lines using supervised and unsupervised hierarchical clusterings. **P,W** (months 3-8).

>>>This Aim was pursed, however, we determined that it was not technically feasible. Namely when fMaSC/fStromal samples were collected, there was not enough RNA present to run a Perou Lab Whole Genome Custom microarray, which requires >1ug of total RNA. Instead we used these precious RNA samples for mRNAseq and single cell RNAseq as discussed below in Aim 2.

1d. The association of the MaSC signature with pathological complete response will be evaluated across multiple data sets with annotated clinical data and where gene expression microarrays had been performed in the pre-treatment samples. Each sample will be assigned an enrichment score as described above. The association of the score with pathCR will be evaluated in all patients and also within each intrinsic molecular subtype as determined by the PAM50-subtype predictor. For those data sets with

survival data, association with survival outcomes will also be evaluated using univariate and multivariate Cox-model analyses. Finally, the enrichment scores during and after chemotherapy will be calculated in the samples of the ISPY-1 trial and also in one publicly available data set where pre- and post-treatment samples after single agent docetaxel or endocrine therapy were profiled. **P** (months 5-8).

>>>>> We accomplished these goals, however, this SubAim required much more work than originally expected; given its importance, this SubAim received more attention and resources than originally described, and did produce a published manuscript (Pfefferle et al., BCRT, 2015 (PMID:25575446). Specifically, we tested the original fMaSC signature, and performed a "refinement approach", and tested multiple fMaSC signatures and determination of their abilities to predict chemotherapeutic response. As mentioned in this Work Statement and our Progress Reports, this Aim included many computational analyses of existing databases in order to explore the prognostic and predictive potential of the fMaSC signature. As originally proposed, we have been reanalyzing the original fMaSC genomic data to "refine" the fMaSC signature. Here, "refinement" means; 1) biological dissection of the fMaSC signature into sub-signatures, and 2) gene set reduction for translation to other technologies. Using a newly derived fMaSC signature coming from a supervised analysis of the fMaSC FACS fraction versus the fStromal+adultMaSC FACS fractions, we identified genes whose high expression better defines fMaSCs as a class of cells. Next we used this ~400 gene set to cluster 300 human breast tumors and determined that the fMaSC signature actually splits into 3 different sub-clusters; one sub-cluster is highest in basal-like tumors, another is highest in luminal tumors, and a third shows no subtype association (Figure 1). This ability to subdivide the fMaSC gene set hints at fMaSC multi-cellular differentiation potential since this single original signature can be broken into distinct smaller signatures that track different cell lineages.

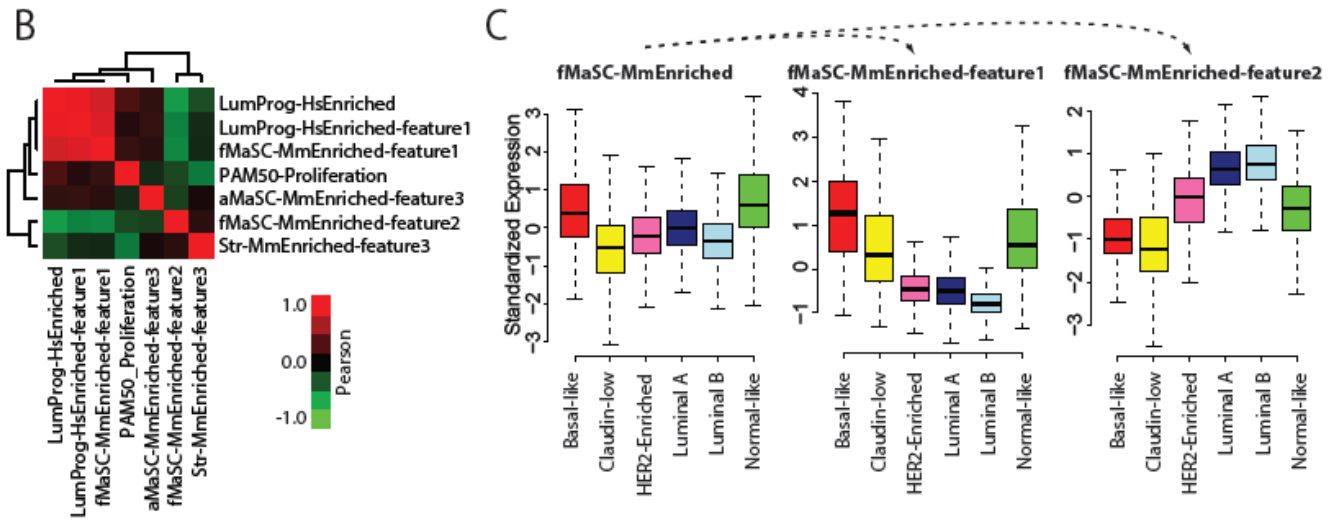


**Figure 1. fMaSC signature and its comparison to other signature and to intrinsic subtypes. In Panel B multiple normal breast mammary cell FAC sorted populations are compared to each other, which shows the relatedness between the fMaSC and Luminal Progenitor signatures. In Panel C the "refined" fMaSC signature components are analyzed for associations with intrinsic subtype where it is seen that the fMaSC-1 is highest in Basal-like and the fMaSC-2 is highest in Luminals.**

We next explored the clinical potential of the three fMaSC sub-signatures using 480 tumors taken from the public domain, including ISPY samples as we had originally proposed. These tumors all came from patients treated with anthracycline and taxane containing neoadjuvant chemotherapy regimens. Even after accounting for the usual clinical and genomic that have been used to predict a likelihood of pCR, the complete fMaSC signature proved to be a significant response predictor (Figure 2).

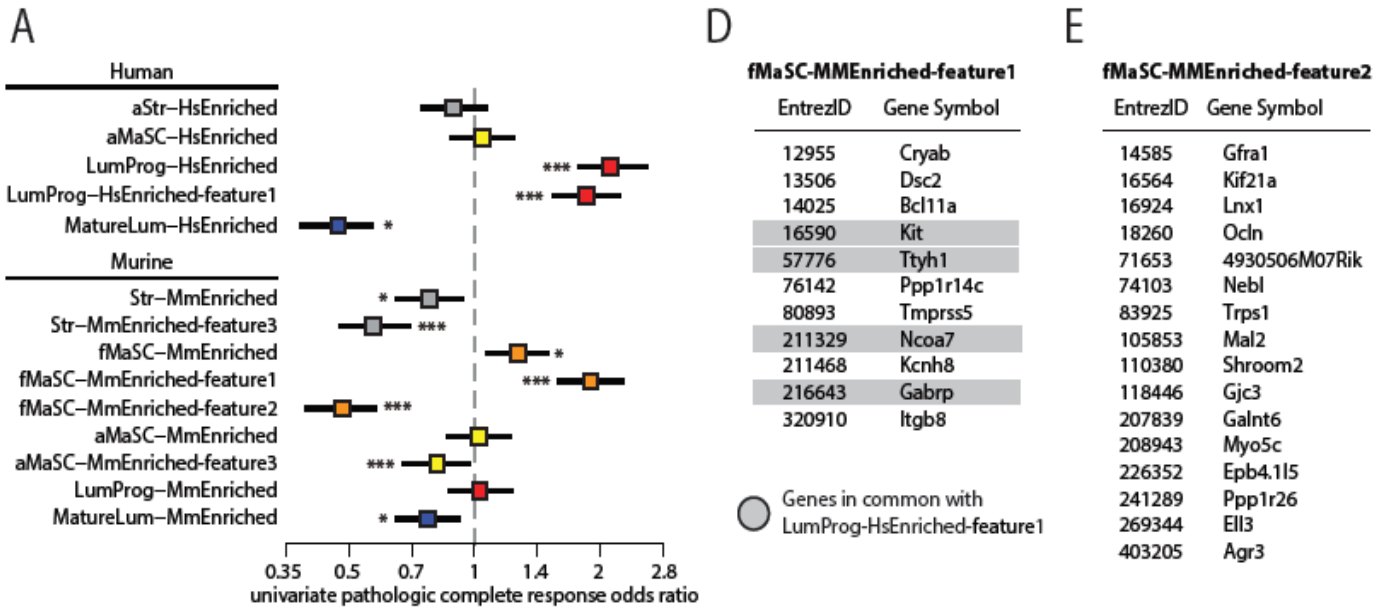


**Figure 2. Clinical significance of the fMaSC and other mammary epithelial cell signatures. A) Forest plot of Odds Ratios for predicting pCR according to multiple different human and mouse epithelial cell signatures. Note that the Luminal Progenitor and fMaSC signatures each predicted a higher likelihood of response, while stromal and luminal features predicted a lower response rate. D) The list of genes in the fMaSC-enriched-1 signature, and E) list of genes for the fMaSC-enriched-2 signature.**

In addition, each of the fMaSC sub-signatures proved to be significant predictors; the fMaSC-basal-enriched signature predicted chemo-sensitivity, while the fMaSC-luminal-enriched signature predicted chemo-resistance. Thus we achieved the overall goals of this Aim, which was to "refine" the fMaSC signature and determine if it was of prognostic and/or predictive abilities on human tumors, and it proved to be predictive of response to neoadjuvant chemotherapy. These results were published in Pfefferle, Spike, Wahl, and Perou (2015) Breast Cancer Res Treat (PMID:25575446).

Ie. The association of the MaSC signature with the development of distance metastases will be evaluated using univariate and multivariate analyses in >800 primary tumor samples were the location and time of the first site of distant relapse was documented (within all patients and also each of the intrinsic subtypes). A signature enrichment score will be calculated as described above. A similar approach will be done for those samples of the UNC database where match primary and metastatic lesions were profiled. **P** (months 7-10).

>>>This has been accomplished using the database of >800 samples coming from Harrell et al., BCRT 2011 (PMID: 21671017); we determined that the fMaSC signature(s) were not associated with specific sites of metastasis. The fMaSC signature(s) remained prognostic when testing for relapse (complete fMaSC signature but not the refined-1 or refined-2), but none of these signatures were prognostic of any one site of metastasis except for the complete fMaSC signature predicting a lower likelihood of metastasis in the bone.

If. The genes comprising refined fMaSC and fStromal signatures will be prioritized according to their novelty as potential therapeutic targets and tractability for functional testing. Scientific/Clinical literature will be surveyed to determine novelty. Scientific/Clinical and Company literature will be surveyed to determine the availability of reagents for functional testing. Cell line expression profiles from the ATCC breast cancer collection and the lines referenced in task Ia. will be bioinformatically evaluated for the presence of signatures suggesting relevance of fMaSC and fStromal genes in the prioritized list and these cell lines will be selected for functional analysis. **P,W** (months 2-l 0)

>>>This has been accomplished and we have identified a number of cell lines that are enriched for the fMaSC signature (BT474 and SKBR3), and we have identified two "refined" gene sets for the fMaSC signature (see Figure 2 above, lists called fMaSC-enriched-1 and fMaSC-enriched-2). Surprisingly, the fMaSC signature is a complex signature that shows both Basal-like and Luminal-like features, and our unique analysis method allowed us to deconvolute this signature into these two components.

lg. Reagents such as small molecule inhibitors, receptor specific antibodies and gene clones or inhibitory RNA constructs will be collected for the top candidates. Activating and deactivating genes will be cloned into existing inducible lentiviral vectors. High titer lentivirus will be produced and validated for inducibility and RNA inhibition or protein production will be validated as appropriate. Other targeted reagents will be validated using standard molecular biological approaches where necessary. **W** (Months 8-14)

>>>This subaim was not fully pursued as we felt the opportunities to focus on the computational analyses were more promising than to pursue the functional characterization of a few genes using cell line models; in addition, we also focused on the single cell RNAseq experiments (see Aim 2) from a computational perspective as well. We made this choice for two reasons. First, the fMaSC signature(s) showed prognostic value AND predictive value for chemotherapy benefit; thus we felt compelled to further pursue this very promising potential biomarker approach. Second, once we discovered that the fMaSC signature had at least two components, we realized that two simultaneous experimental paths would have to be taken, one for the fMaSC-refined-1 gene set (basal-like), and a second for the fMaSC-refined-2 (luminal) gene sets. Instead of spreading ourselves too thin on these in vitro studies, we decided to focus more on the computational analyses, which were extensive and included the analysis of numerous data sets including multiple human breast cancer sets, and the analysis of the single cell RNAseq data. Our computational studies did identify a potential new biomarker of chemotherapy sensitivity in human patients, which is being tested in clinical studies including CALGB 40603.

1h. The ability of viral vectors or other reagents to specifically impact their molecular targets will be examined in breast cancer cell lines through 2D and 3D culturing systems and standard molecular biological approaches (e.g. western blotting, immunofluorescent/cytochemistry, RT-PCR etc. examining receptor phosphorylation, protein localization, mRNA abundance, etc.). Cellular effects on proliferation, survival and migration will also be assayed. **W** (months 14-18)

>>>This subaim was not pursued as we felt the opportunities to focus on the computational analyses were more promising (see Aim 1g for rationale)

1i. Cell lines exhibiting biological responses to activation or inhibition of fMaSC and fStromal pathways in vitro will be transduced with lentiviral vectors and injected as xenografts into immune compromised mice. Tumor growth and metastasis will be evaluated in real time using luminescent imaging. **W** (months 16-24)

>>>This subaim was not pursued as we felt the opportunities to focus on the computational analyses were more promising (see Aim 1g for rationale)

**Task 2. Embryonic mammary stem cell signature refinement using RNA-seq and functional validation**

2a. The Wahl lab will obtain timed pregnant female mice, obtain embryos from El8.5, dissect mammary rudiments, isolate fMaSC and fStroma by methods they have developed. The cells will be flow sorted to obtain fMaSC enriched (CD49fhighCD24highNCAM-) and fStromal (CD49f+, CD24-/+) populations. **W** (months 1-2)

>>>This has been accomplished (see data below).

The Lasken lab will receive fMaSC enriched and fStromal cells from the Wahl lab. The fMaSC enriched cells will be micromanipulated to obtain single cells, which will then be lysed to preserve RNA integrity and maximize efficiency for generating eDNA for SOLiD sequencing. **L** (months 1-2).

>>>>This was accomplished (see data below).

2b. cDNA will be generated from each cell and amplified by a PCR method under conditions suitable for subsequent SOLiD DNA sequencing **L** (months 2-3).

2c. Each single cell cDNA preparation will be pre-screened for the fMaSC phenotype based on expression of Kl4 and K8 using qRT-PCR **L,W** (months 2-3).

2d. The highest quality eDNA samples representing 10 additional K14+K8+, 5 K14+K8- and 5 Kl4-K8+ cells will be RNA sequenced by the SOLiD sequencing method at a level generating about 40 million sequence reads/cell. **L** (months 3-6).

>>>>Tasks 2a-2d were initially done manually using a variety of methods. However, approximately 18 months ago, a Fluidigm C1 single cell microfluidic instrument capable of lysing 96 single cells in situ, and then preparing cDNA samples from them was obtained by the California Institute for Regenerative Medicine (CIRM), which is next to the Salk. We were granted approval to use this instrument. We isolated approximately 100 individual cells from e18.5 in several independent experiments and loaded them onto the C1 to obtain cDNA. An example of the data obtained is shown in Figure 1. Of note, this instrument enables one to evaluate the cells microscopically using a live-dead cell dye to restrict all RNA-sequencing data to live cells. As can be seen, we detected cells co-expressing K14 and K8 RNA (Figure 1, Cells #4-6), and the RNA seq data follow the gene models very well.

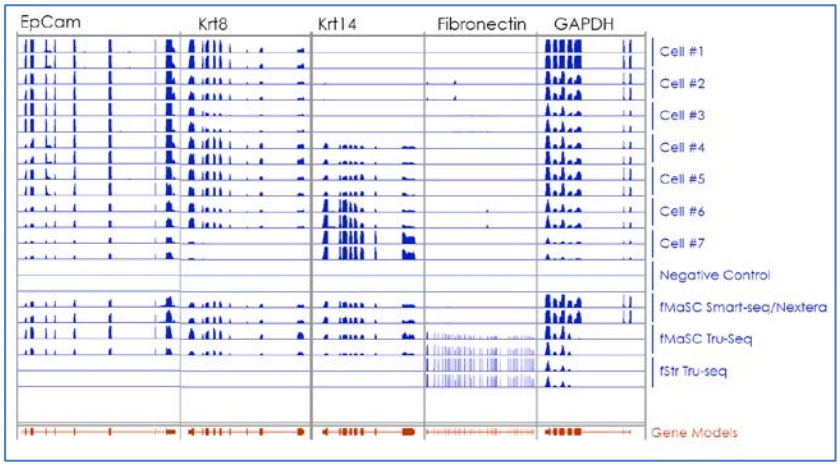


Figure 1. Genomic sequence alignments of RNA derived sequences from several individual fMaSC cells and control samples. The alignments show high concordance with annotated exon structures (Gene Models). The data also show high technical reproducibility between replicate sequencing experiments for each sample. Controls are comprised of pools of fMaSC cells processed using the same biochemistry as the C1 protocol (fMaSC Smart-seq/Nextera) or lacking reverse transcriptase (Negative control). An fMaSC pool and a Stromal pool (fStr) processed by an alternative approach (Tru-Seq) that works on bulk samples are also provided for comparison.

2e. RNA-Seq data will be analyzed to discover additional genes and gene clusters associated with the fMaSC cells. These data will be combined with the analysis of 10 Kl4+K8+ cells that are currently be sequenced with funding from the JCVI and a Salk Cancer Center Starter award to yield a total of 20

Kl4+K8+ cell analyzed. The sequence and data analyses will be conducted jointly by the Lasken and Wahl labs. **L, W** (months 7-13).

2f. A list of markers will be generated through bioinformatics analysis of single cell RNA-Seq data to identify markers associated with distinct cell types. The literature will be surveyed for the availability of reagents for the prospective isolation of the distinct cell types using the identified markers. Reagents will be acquired. Cells will be isolated based on these markers and the fidelity of separation of the individual cell types and per based analysis resorting will be carried out to assess purity of sorting. **P, L, W** (months 12-15)

>>>>Our initial evaluation of the gene expression profiles within the fMaSC population at single cell resolution showed the cells to be heterogeneous with no clear subpopulation likely to correspond to a distinct stem cell subpopulation. However, through the use of the C1 instrument, we were able to broaden our research approach to include additional developmental states that could be used to delineate gene expression changes that define the gain and loss of the stem cell phenotype over the course of development. That is, instead of focusing just on E18 cells, we decided to obtain cells from throughout development so that we could have a data set that would position us to identify the pathways that are altered in going from the pre-stem cell state at e16, to the stem cell state at e18, and then into the differentiated myoepithelial and luminal lineages associated with adult mammary development. We have now sequenced hundreds of cells from E18, P0, P4 and adult mice, and have clustered the data using multiple bioinformatic methods to assign cellular phenotypes. As one example, we used the Monocle strategy to infer lineages based on generation of minimum spanning trees of transcriptional relatedness (Figure 2 A,B). We then derived an independent approach that does not use the Monocle assumption of direct lineage relationships between cells of close transcriptional relatedness, which resulted in a very similar outcome (Figure 2 C-E)

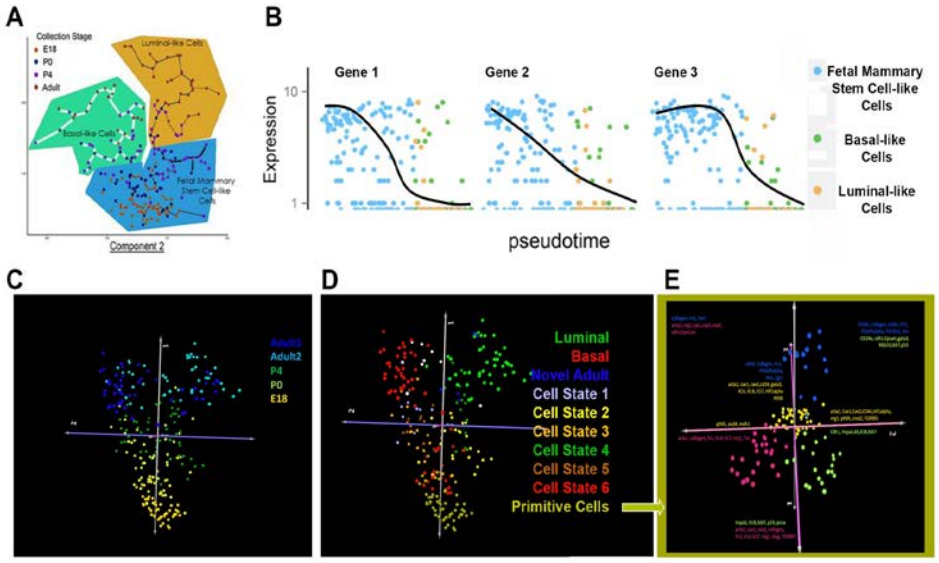


**Figure 2. Unsupervised identification of cell types and candidate regulatory genes from single cell data.** A) Monocle plot of single cell relatedness (i.e. proximity in the graph), minimum spanning tree model of differentiation ('pseudotime') and identification of three putative cell types (Luminal-like, Basal-like and fMaSC-like). B) Expression levels of three highly fMaSC associated genes in single cells that have been reorganized according to their position along the pseudotime, minimum spanning tree. Note: The majority of adult cells are plotted on the x-axis for these three genes as they are rarely expressed in adult cells (i.e. y=0). C-D) An alternative approach yields similar but more refined results. C) Using our alternative approach cells are found to be distributed along a continuum from E18 to adult (vertical axis). Clustering of these cells according to gene expression ranks, identifies known Luminal and Basal adult cell types, a novel adult cell type and several cell states along the continuum from the most primitive cells to the adult. D) Fine scale analysis of the most primitive cells identifies genes associated with the earliest differentiation events as bi-phenotypic cells become more luminal (green), more basal (pink) or more niche related (blue).

These methods proved robust for separating cells according to the developmental time at which they were obtained, and to the lineages to which they are most related. Further analysis based on these approaches has enabled us to define a subset of cells from E18 that appear to be uncommitted to either the luminal or myoepithelial lineages, and yet to express genes indicative of both. For example, these cells express both luminal and myoepithelial

cytokeratins, as well as lineage specification genes including Sox10, GATA3, and Elf5. The methods and examples of data obtained from these analyses were presented in the Progress Report submitted last year. A manuscript describing these studies is currently being prepared and is only awaiting data obtained from cells in the "pre-" stem cell state at E15-16.

We have experienced significant technical problems using the C1 instrument to analyze cells derived from E15-16. Recently, we have returned to working with the Lasken lab to sort cells from E15-16 cells directly into wells of a 384 well plate, and then used the SMART-seq 2 protocol to prepare cDNA libraries. Sequencing of these libraries is currently in progress.

2g. Cells will be sorted using population specific markers. In vitro colony growth, serial replating ability and immunofluorescent analysis ofbipotent progeny will be evaluated for each candidate marker. **W** (month 13-18)

2h. Markers yielding stem cell phenotypes in vitro will be used to sort fetal mammary cells. Cells will be transplanted at limiting dilution to reconstitute murine mammary glands additionally single cells will be transplanted to reconstitute murine mammary glands. **W** (months 18-24)

2i. Additional RNA-seq will be performed to refine the data obtained from validated fMaSCs. A second SOLiD sequencing run will be carried out on eDNA from ten single cells to refine and test conclusions obtained in the first year of the grant. **L,W** (Months 14-18).

>>>>We have accomplished the major goals of Aims 2g-i, and provide the following as one important example of the value and relevance of the data obtained. The results, summarized briefly below, identified the cell state regulator SOX10 as a developmental control gene able to identify and highly enrich for fMaSCs that is also required for the fMaSC state (Dravis et al (2015), Cell Reports, v. 12, pgs 2035-2048).

Our expression profiling identified SOX10 as one of the most differentially regulated genes in fMaSCs using both microarrays and single cell RNA sequencing (Dravis et al, Figure 1A, pg 2036). We obtained a mouse expressing an H2B-Venus transgene under the control of the SOX10 endogenous promoter. The mammary epithelial cells in the embryonic rudiments were brilliantly labeled, while there was little if any staining in the surrounding stroma (Dravis et al, Figure 2A, pg. 2038). We used FACS to obtain various cell fractions on the basis of their expression of different levels of EpCAM or SOX10 (i.e., venus fluorescence). We found that only those cells that were EpCAM+ and SOX10-high exhibited all of the properties expected of fMaSCs: generation of polarized organoids in vitro, ability to generate full functional mammary outgrowths from limited numbers of cells transplanted into de-epithelialized fat pads, and ability to "self-renew" as assayed by multiple rounds of transplantation, or dissociation and re-formation of spheres in vitro (Dravis et al, Figure 2E, F, G, H, I, pg 2038). We also obtained mice with floxed SOX10 genes, deleted the SOX10 genes from fMaSCs in vitro, and found that fMaSCs lacking SOX10 expression no longer formed organoids in vitro or transplanted in vivo (Dravis et al, Figure 4A-E, pg. 2041). Finally we showed that over-expressing SOX10 led to two very important phenotypes. First, after short periods of expression, we found the fMaSCs could form secondary organoids with much higher efficiency than if they did not express high SOX10 levels. Second, if SOX10 expression was maintained at high levels, the fMaSCs lost expression of epithelial markers, no longer expressed luminal or basal cytokeratins, gained expression of vimentin, and became motile but non-proliferative (Dravis et al, Figure 5A-F, pg. 2042). In other words, they acquired many characteristics of cells that had undergone an epithelial-mesenchymal transition. Importantly, reducing SOX10 expression in the cells that had moved away from the organoids to set up solitary "satellites" resulted in reversion of the cells to an epithelial state, re-entry into the cell cycle, and restoration of their ability to generate both luminal and myoepithelial descendants. In other words, the stem state was readily reversed depending on SOX10 levels.

We have begun to search for in vivo conditions that regulate SOX10 in the mammary gland and that could be relevant to fMaSC genesis and breast cancer biology. We found that FGF10 specifically induces SOX10 transcription, and that either leaving SOX10 out of the culture medium, or using an pan-FGF receptor inhibitor, prevents SOX10 transcriptional activation, and prevents fMaSCs from undergoing an EMT (Dravis et al, 2015, Figures 1B-E, pg 2036). Interestingly, FGF10 is one of the factors produced during wound healing. We speculate that as wound signatures have been correlated with initiation and progression of breast cancer, that exposure of the fMaSC-like cells we have documented to be present in basal-like breast cancers may enable them to acquire motility, depart the local tumor environment, and metastasize to distant sites at which, if exposed to a lower FGF environment, then may reverse their phenotype, become more stem-like, and produce a heterogeneous cellular mass at an ectopic location.

2j. Identification of gene signatures corresponding to fMaSC from bioinformatic analysis (task 2e,f ) and bioinformatic refinement/reduction of the signature. Selection of candidate markers for analysis of fMaSC contribution to archival tumor samples and tissue analysis **P,L,W** (months 12-24)

>>>>We found that elevated Sox10 expression is found in Basal-like and some Claudin-low human breast cancers (Dravis et al, 2015, Figure 1F, pg 2036).

2k. Immuno-histochemical and in situ analysis of archival tumor tissue. **P,W** (months 12-24).

>>>>We are now developing the collaborations we need to obtain relevant samples from UCSD, and we continue to work with Dr. Perou to analyze his human and mouse tissue samples as we derive additional informative signatures. Unfortunately, we have found no Sox10 antibodies suitable for IHC or IF analyses.

**Key Research Accomplishments:**

Aim 1

1. Development of a meta-analysis approach to derive more precise signatures for normal mammary cell luminal, progenitor, myoepithelial, and stem cell populations from human and mouse systems. This method proved more robust than using single studies for analysis, and sets a precedent for use of such meta-analysis-derived signatures in future studies.
2. Application of refined signatures based on normal mammary cell types to analysis of human breast cancers and mouse cancer models to determine which normal cell types correspond most closely to cancers in each species.
3. Use of single sample classifiers revealed diversity of cellular relationships among each GEMM and human breast cancer intrinsic subtype.
4. Demonstration that the human luminal progenitor and one feature of the mouse fMaSC signature correlate with pCR across all human breast cancer subtypes, and retains significance in multi-variable analyses including proliferation, subtype, and clinical parameters. Importantly, one feature of the fMaSC profile associated with luminal attributes predicted for poor response to anthracycline/taxane based chemotherapy for patients whose tumors display enrichment for this profile.

Aim 2

1. Obtained transcriptomes from hundreds of individual cells across four developmental time points critical for understanding mechanisms of acquisition and loss of the stem cell state during mouse mammary development.
2. Use of transcriptomic data to identify candidate transcriptional regulators relevant to acquisition of mammary stemness. Identification of SOX10 as one such gene.
3. Demonstrated fetal mammary cells expressing SOX10 uniquely identify the stem cell population. This discovery facilitated purification of the most pure fMaSC population to date, which enabled obtaining more precise transcriptomic data.
4. Genetic strategies were employed to show that SOX10 is required for fMaSC function in vitro and in vivo.
5. Developed a genetic system to enable analysis of the effects of SOX10 overexpression. These studies showed that persistent SOX10 expression preserves fMaSC multipotentiality, but long term high SOX10 expression causes fMaSCs to undergo a mesenchymal transistion that does not correlate with elevated levels of Slug, Snail, Zeb1, or Twist as reported for other systems. The mesenchymal transition was reversed upon reducing SOX10 levels.
6. Gene expression and functional studies revealed a positive feedback loop between FGF signaling and SOX10. Elevated SOX10 led to upregulation of potentiators of FGF signaling, and down regulation of FGF signaling antagonists.

**Conclusion**

Our new data are consistent with our previous studies showing that fMaSC signatures contain unique combinations of expressed genes with relevance to human breast cancer biology, including the response of breast cancers of all intrinsic subtypes to chemotherapy. We have thus developed a potentially useful metric for clinical decision making. We continue to improve methods for doing single cell RNA-seq, and for bioinformatically analyzing the results. These studies revealed the potential relevance of SOX10 to fMaSC biology, which we established using a combination of in vitro and in vivo approaches.

**Publications:**


1. Pfefferle, A.D., Herschkowitz, J.I., Usary, J., Harrell, J.C., Spike, B.T., Adams, J.R., Torres-Arzayus, M.I., Brown, M., Egan, S.E., Wahl, G.M., Rosen, J.M., and Perou, C.M. (2013) Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. Genome Biol. 14(11):R125 epub ahead of print. PMCID:PMC4053990
2. Dravis, C., Spike, B.T., Harrell, C.J., Johns, C., Trejo, C., Southard-Smith, M.E., Perou, C.M. and Wahl, G.M. (2015) Sox10 regulates stem cell and mesenchymal states in mammary epithelial cells. Cell reports, *pii: S2211-1247(15)00925-0.* NIHMSID722490


**Abstracts and Presentations:**

**Geoff Wahl, Ph. D.**

1/1/15: Speaker, Moores Cancer Center's Head & Neck Cancer Retreat, UCSD

4/10/15: Keynote Lecture, University of Arizona Cancer Center

5/15/15: Speaker, University of Chicago Seminar Series, University of Chicago

**Chris Dravis, Ph.D.**

4/22/15 AACR Annual Meeting, Philadelphia, PA (Poster)

**Benjamin T. Spike, Ph.D.**

1/12/15 University of Utah, Huntsman Cancer Institute

4/2/15 University of California, San Diego, Department of Pathology

4/30/15 University of Colorado, Denver, Department of Pathology

**"Inventions, Patents and Licenses"**

None.

**Reportable Outcomes**

- fMaSC gene signatures correlated to chemotherapeutic response
- Sequencing and analytical pipeline for single cell RNA Sequencing
- SOX10 as a marker and functionally relevant transcription factor contributing to the mammary stem cell state
- Involvement of FGF signaling in SOX10 induction, and regulation of stem and mesenchymal states in the mammary gland

**Other Achievements**

G.M. Wahl

- Successfully competed for Outstanding Investigator Award funding: 7 years, ~$600,00 per year

B.T. Spike

- Obtained Assistant Professor position at University of Utah

**Appendices**

None